MARKOV CHAINS OF ORDER GREATER THAN ONE*

WILLIAM P. LOWRY and DONALD GUTHRIE
Department of Statistics, Oregon State University, Corvallis, Oreg.

ABSTRACT

This paper proposes means whereby a test for the goodness of fit and an estimate of the order of a Markov chain model may be obtained concurrently. Application of the method to daily precipitation data for two seasons of about 30 yr. in New Mexico, Colorado, and Oregon suggests two tentative conclusions. First, within a single climatic area, order estimates tend toward zero as precipitation events become rarer. This may occur at the drier stations or at higher thresholds defining a wet day. Second, between climatic areas, the station with the greater diversity of air mass types will tend to have order estimates greater than the station having the same seasonal mean precipitation but a less diverse climate.

1. INTRODUCTION

In 1924, Besson [2] reached the conclusion, through a statistical analysis, that at Montsouris, France, past weather exerts an influence on future weather. In their well-known monograph Brooks and Carruthers ([4], p. 315) provide a statistical test for the presence of persistence in sequences of weather events. The notion of persistence in weather and climate is certainly not new, but only recently have the Markovian models been applied to climatological processes exhibiting sequential patterning.

Papers using Markovian models in recent years have met with considerable success. Most have dealt exclusively with occurrences of wet and dry days. Typical of those employing discrete time and stationary transition structure are the familiar papers by Gabriel and Neumann [9], Caskey [5], and Hopkins and Robillard [13]. Green considered continuous time [10], Feyerherm and Bark [6, 7] dealt with transition structure changing smoothly through the year, and Weiss [14] included mention that sequences of weather types described in more elaborate terms than simply "wet" and "dry" are amenable to such treatment.

Several of these studies have employed no statistical tests for goodness of fit between models and observations. Only two (Gabriel and Neumann; Feyerherm and Bark) appear to have considered orders of dependence greater than one in their Markov chain models.

While noting also that none of these papers has considered models having more than two states, it is also worth noting the data which would be required to do so. Since the number of state-sequences required to define transition in a t^{th}-order s-state process is s^{t+1}, one may expect a drop in the mean cell frequency with a small increase in s such as to make weather records at most stations (25–40 yr.) too short for consideration of time units the order of a month. For example, a 30-yr. record considered by 30-day months would have 900 counts to distribute among 4^5, or 1,024, cells in a model of four states and order 4.

In this paper we wish to present a method and some results wherein a) stationarity is assumed, b) testing the fit of models and observations takes place, and c) the order of the Markov chain is estimated for those processes which exhibit Markovian characteristics. The method allows as well for increasing the number of states in the model without undue difficulty.

2. METHODS

Following the theory of Billingsley [3] and the recommendations of Hoel [12], we propose the following method for various types of modelling and testing Markov chains in climatology.

The events in a sequence are restricted to a finite list of possible outcomes, or states, $a_1, a_2, \ldots, a_s$. The process, $\{x_n\}$, represented by observed sequences is an s-state process. If the probability that the process will enter state a_k on the kth step depends upon the sequence of states in steps $(k-t)$ through $(k-1)$, and on no other, then the process is a tth-order s-state process.

In most of the papers referred to above the chains modelled were simple chains with a maximum allowable value of $t=1$: first order. In several, as noted, models with $t=2$ were considered. In our approach, the conditional probabilities governing entrance into the various states at each step are of the form:

$$p_{a_1 \ldots a_t : a_{t+1}} = \Pr\{x_n = a_{t+1} | x_{n-t} = a_1, \ldots, x_{n-1} = a_t\}.$$

Each of these is termed a transition probability, and the matrix of all these is the transition structure of the process. In a sample of the process sequence, the observed frequency of the n-step transition from state a_1 through

*Research supported in part by the Atmospheric Sciences Section, National Science Foundation, NSF Grant GA-379.

states $a_2, \ldots, a_{n-1}$ and to state a_n is denoted by $f_{a_1 \ldots a_{n-1}:a_n}$. There is a separate frequency, f, for each of the s^n possible permutations of the states.

The question of whether or not the observed sequence represents a tth-order Markov chain is answered by application of a Chi-square test with the statistic:

$$\chi^2 = \sum_{a_1 \ldots a_{t+1}} \frac{(f_{a_1 \ldots a_{t+1}} - f_{a_1 \ldots a_t} p_{a_1 \ldots a_t : a_{t+1}})^2}{f_{a_1 \ldots a_t} P_{a_1 \ldots a_t : a_{t+1}}}$$

and $s^{t+1} - s^t$ degrees of freedom.

To make use of the test, of course, one must have not only the counts, or frequencies involved, but also an estimate of the transition probability, p. Billingsley [3] and Hoel [12] recommend assuming that the process being analyzed is of some high but manageable order, t, and then testing the $(r+1)$ hypotheses that the chain is of order $0, 1, 2, \ldots, r$ $(r<t)$ within the higher order assumption. The test statistic resulting from these assumptions and hypotheses is:

$$\chi^2 = \sum_{a_1 \ldots a_{t+1}} \frac{\left[f_{a_1 \ldots a_{t+1}} - f_{a_1 \ldots a_t} \left(\frac{f_{a_{t-r+1} \ldots a_{t+1}}}{f_{a_{t-r+1} \ldots a_t}} \right) \right]^2}{f_{a_1 \ldots a_t} \left(\frac{f_{a_{t-r+1} \ldots a_{t+1}}}{f_{a_{t-r+1} \ldots a_t}} \right)}$$

for the hypothesis that the order is r; for $r=0$,

$$\chi^2 = \sum \frac{\left[f_{a_1 \ldots a_{t+1}} - f_{a \ldots a_t} \left(\frac{f_{a_{t+1}}}{N} \right) \right]^2}{f_{a_1 \ldots a_t} \left(\frac{f_{a_{t+1}}}{N} \right)}$$

where N is the total number of transitions counted. As the total number of transitions becomes large, the statistic is asymptotically distributed as Chi-square with $(s^{t+1}-s^t)-(s^{r+1}-s^r)$ degrees of freedom, under the null hypothesis. The estimate of the order of the process is taken to be the smallest value of r which produces a nonsignificant test statistic. In this way the estimating scheme also provides the test for goodness of fit.

The results to be described in this paper are from models in which $s=2$, $t=4$, and $r=0,1,2$, and 3. The computer program calculates from a sample sequence the counts of the 2^5, or 32, possible permutations of the two states wet (W) and dry (D) taken five at a time. As an example, in one case it was found that the sequence *DDWWD* was observed 13 times, out of the 36 times which *DDWW* was observed. Under the full 4th order model, the estimate of the transition probability p_{ddwwd} is 13/36. To test the assumption that the process is of order 2 rather than 4, we observe that the sequence *WWD* occurred 50 times out of the 101 times that the sequence *WW* was observed. Hence, the contribution to the test statistic (a sum taken over 32 such sequences) of this particular sequence was:

$$\Delta_{ddwwd} = [13 - 36(50/101)]^2/(36)(50/101) = 1.303.$$

For the test of the hypothesis that $r=2$, the degrees of freedom used are $(2^5-2^4)-(2^3-2^2)$, or 12.

3. RESULTS

In 1965 the methods described were applied to the daily precipitation records of two Oregon stations [1]. They are Seaside, a coastal station with wet winters and relatively dry summers, and Squaw Butte, a station in the high desert where both winter and summer are dry by comparison with the coastal areas. The Seaside data were for the 33 yr. 1931–1963, and those for Squaw Butte were for the 28 yr. 1937–1964. The records were each examined for a winter period and a summer period with a threshold dividing the two states taken to be 0.01 in. of precipitation or more as a wet day. Three definitions were used for each of the two seasonal periods, and Seaside's winter periods were analyzed with an additional threshold of 0.20 in. of precipitation. The results are summarized in table 1, where both the definitions of the seasonal periods and the Chi-square statistics are shown.

Subsequent to the publication of Oregon's methods and results, Heermann [11] at Fort Collins and Finkner [8] at Las Cruces undertook similar tests for Colorado and New Mexico data as part of the research program of the Western Regional Research Project, W-48, of the U.S. Department of Agriculture. Heermann and Finkner used more stations and more thresholds than were used in Oregon, the same values for t and r, but fewer definitions of "winter" and "summer." Their results are summarized in tables 2 and 3.

4. DISCUSSION

When the order estimates from tables 1, 2, and 3 are combined as in table 4, and when these estimates are then used to compute mean order estimates for each threshold, as in table 5, the tendency is readily apparent that the order of the Markov chain decreases as the threshold defining states increases. This tendency is due to the increase in relative rarity of one of the states (W) as the

TABLE 1.—*Summary of Chi-square statistics for two Oregon stations*

Station	Period	Threshold (in.)	Test value of r				Order estimated ($P<0.05$)
			0	1	2	3	
Seaside	Jan	0.01	175	8.54	8.19	6.26	1
	Feb	.01	172	27.0	22.4	6.65	3
	Jan. 15–Feb. 20	.01	219	19.8	16.6	7.64	1
	Jan	.20	150	27.2	9.38	1.94	2
	Feb	.20	112	25.9	22.8	8.13	3
	Jan. 15–Feb. 20	.20	182	30.0	18.1	6.87	2
	July	.01	80.7	15.2	9.54	4.11	1
	Aug	.01	66.0	20.2	14.5	8.88	1
	July 10–Aug. 15	.01	69.1	25.5	16.5	13.5	2
Squaw Butte	Jan	.01	64.2	33.5	23.8	18.3	>3
	Feb	.01	168	31.9	30.3	19.0	>3
	Jan. 15–Feb. 20	.01	67.8	29.9	21.1	11.6	2
	July	.01	70.2	27.6	12.7	7.50	2
	Aug	.01	67.5	15.8	9.51	5.80	1
	July 10–Aug. 15	.01	58.3	26.4	15.1	9.10	2

TABLE 2.—(a) *Selected Chi-square statistics for New Mexico stations and (b) summary of order estimates for 14 New Mexico stations*

Station (a)	Period	Threshold (in.)	Test value of r: 0	1	2	3	Order estimated ($P<0.05$)
Bloomfield	Jan	0.01	69.1	35.1	17.9	13.5	2
		.03	70.6	36.6	16.3	15.2	2
		.05	67.2	37.8	20.0	17.7	2
		.10	51.8	30.9	11.4	4.92	2
		.20	19.9	19.9	18.6	1.01	0
		.30	7.40	4.23	3.92	3.67	0
	July	.01	47.0	15.5	10.7	9.49	1
		.03	54.5	15.2	10.3	6.81	1
		.05	43.3	19.2	11.9	10.6	1
		.10	22.5	10.1	5.88	3.42	0
		.20	10.9	9.04	6.10	1.40	0
		.30	5.15	2.21	1.32	0.63	0
Corona	Jan	.01	53.4	14.5	7.44	6.56	1
		.03	53.7	14.8	7.72	6.72	1
		.05	65.6	32.6	14.7	9.14	2
		.10	46.9	15.5	5.25	3.94	1
		.20	33.5	20.0	11.9	10.0	1
		.30	27.3	20.7	6.60	3.28	1
	July	.01	52.5	19.6	14.5	3.69	1
		.03	49.9	18.7	16.4	4.73	1
		.05	46.4	17.0	15.4	8.40	1
		.10	33.1	8.20	6.47	3.64	1
		.20	19.1	9.71	7.38	4.53	0
		.30	19.1	8.82	8.10	3.44	0

Station (b)	Period	Threshold (in.): 0.01	0.03	0.05	0.10	0.20	0.30
Albuquerque	Jan	1	1	1	0	0	0
	July	1	1	1	1	0	1
Bloomfield	Jan	2	2	2	2	0	0
	July	1	1	1	0	0	0
Cimmaron	Jan	0	1	1	1	1	0
	July	1	0	0	0	0	0
Clayton	Jan	1	1	1	0	0	0
	July	1	1	1	1	1	1
Clovis	Jan	1	1	2	1	1	0
	July	1	0	0	0	1	1
Corona	Jan	1	1	2	1	1	1
	July	1	1	1	1	0	0
Fort Bayard	Jan	1	>3	2	2	1	0
	July	1	1	1	1	1	1
Lovington	Jan	0	0	0	0	0	0
	July	1	1	1	0	0	0
Mosquero	Jan	2	2	3	3	1	0
	July	1	1	1	1	1	1
Roswell	Jan	2	1	1	1	1	0
	July	1	1	1	1	2	1
Santa Fe	Jan	1	1	0	0	0	0
	July	1	0	0	0	0	0
Socorro	Jan	1	1	1	0	0	0
	July	1	1	0	0	0	0
Tucumcari	Jan	1	2	2	0	0	0
	July	1	1	1	1	0	1
University Park.	Jan	>3	1	1	1	0	0
	July	0	0	0	0	0	0

TABLE 3.—*Summary of Chi-square statistics for three Colorado stations*

Station	Period	Threshold (in.)	Test value of r: 0	1	2	3	Order estimated ($P<0.05$)
Burlington	Jan	.01	22.2	9.16	5.41	1.44	0
		.10	10.6	5.11	2.61	0.45	0
		.20	11.4	3.26	2.62	2.33	0
		.50	49.1	48.9	48.6	48.4	>3
	July	.01	24.6	13.5	13.2	8.32	0
		.10	34.6	21.2	18.4	13.5	1
		.20	19.9	16.7	6.73	2.93	0
		.50	45.0	35.4	26.8	10.4	3
Fruita	Jan	.01	72.6	39.3	27.4	18.8	>3
		.02	51.2	23.9	15.8	9.42	2
		.03	38.4	16.4	13.6	6.58	1
		.05	40.3	16.8	11.4	7.00	1
		.10	34.3	20.3	16.5	12.4	1
		.20	6.40	3.35	2.99	1.36	0
		.50	39.0	38.8	0.28	0.26	2
	July	.01	79.0	17.8	17.6	15.3	1
		.02	69.5	16.3	16.5	13.4	1
		.03	54.3	13.4	12.3	9.87	1
		.05	55.4	12.9	9.56	5.00	1
		.10	59.1	10.1	6.35	1.47	1
		.20	38.6	3.35	3.18	1.47	1
		.50	0.18	0.13	0.08	0.04	0
Gunnison	Jan	.01	49.7	12.8	12.5	9.10	1
		.10	39.7	8.30	8.35	7.09	1
		.20	28.7	6.35	2.88	1.89	1
		.50	6.01	5.78	5.57	0.12	0
	July	.01	32.1	21.5	15.8	12.3	1
		.10	9.95	7.07	5.76	3.86	0
		.20	10.7	8.27	4.26	4.28	0
		.50	4.04	2.66	2.23	0.42	0

TABLE 4.—*Summary of order estimates from tables 1, 2, and 3. O= number of Oregon stations; C=number of Colorado stations; N= number of New Mexico stations.*

Period	Threshold (in.)	Order estimated: 0			1			2			3			>3		
		O	C	N	O	C	N	O	C	N	O	C	N	O	C	N
Jan	0.01		1	2	1	1	8			3				1	1	1
	.02								1							
	.03			1		1	9			3						1
	.05			2		1	6			5			1			
	.10		1	6		2	5			2			1			
	.20		2	8		1	6	1								
	.30			13			1									
	.50		1									1			1	
July	.01		1	1	1	2	13	1								
	.02					1										
	.03			4		1	10									
	.05			5		1	9									
	.10		1	7		2	7									
	.20		2	9		1	4			1						
	.30			7			7									
	.50		2									1				

threshold increases. It is further indicated by the presentation of table 6, where the order estimates from the two driest and the two wettest New Mexico stations (as judged by mean seasonal precipitation) show a distinct separation, with wetter records giving higher values of estimated r.

Another interesting point of comparison is between Oregon stations and New Mexico stations having about the same seasonal mean precipitation. In January, Squaw Butte, Oreg., and Corona and Fort Bayard, N. Mex., have comparable mean precipitation values near 1 in. In the comparisons of order estimates between Oregon's station and New Mexico's stations, the result is that at threshold 0.01 in. New Mexico gets estimates of $r=1$ while Oregon gets an estimate that r exceeds 3. Similarly, comparing Squaw Butte with Bloomfield and University Park in July (mean precipitations near ½ in.), the result

TABLE 5.—*Mean order estimates* calculated from table 4*

Period	Stations	Threshold (in.)					
		0.01	0.03	0.05	0.10	0.20	0.30, 0.50**
Jan	New Mexico only	1.29	1.36	1.36	0.86	0.43	0.07
	New Mexico plus Colorado	1.35	1.33	1.33	0.82	0.41	0.47
	New Mexico, Colorado, and Oregon	1.47				0.50	
July	New Mexico only	0.93	0.71	0.64	0.50	0.43	0.50
	New Mexico plus Colorado	0.88	0.73	0.67	0.53	0.41	0.59
	New Mexico, Colorado, and Oregon	0.95					

*The mean order estimate was calculated as $(\Sigma r F_r / \Sigma F_r)$ where r is the estimated order, F_r is the number of stations for which the order estimate is r, and the summations are taken over r.

**Since New Mexico used a threshold of 0.30 and not 0.50, and Colorado used 0.50 and not 0.30, these two results are combined as representative simply of a "large threshold." To separate them would require ΣF_r to be only 3 for Colorado's threshold of 0.50—a small number for these purposes.

TABLE 6.—*Order estimates for the two seasonally driest and the two seasonally wettest New Mexico stations.* $(\underline{N})$ *=the number of dry stations;* $(\overline{N})$ *=the number of wet stations.*

Threshold	January Order estimated					July Order estimated				
	0	1	2	3	>3	0	1	2	3	>3
0.01		$\underline{2}, \overline{2}$				$\underline{1}$	$\underline{1}, \overline{2}$			
.03		$\underline{2}, \overline{1}$			$\overline{1}$	$\underline{1}, \overline{1}$	$\underline{1}, \overline{1}$			
.05		$\underline{2}$	$\overline{2}$			$\underline{1}, \overline{1}$	$\underline{1}, \overline{1}$			
.10	$\underline{2}$	$\overline{1}$	$\overline{1}$			$\underline{2}, \overline{1}$	$\overline{1}$			
.20	$\underline{2}$	$\overline{2}$				$\underline{2}$	$\overline{2}$			
.30	$\underline{2}, \overline{1}$	$\overline{1}$				$\underline{2}$	$\overline{2}$			

is that New Mexico gets estimates of 0 and 1 while Oregon gets an estimate of 2.

A tentative explanation for these larger order estimates in Oregon as compared with similarly wet stations in New Mexico for both winter and summer is that in both seasons the air mass climate in Oregon is more diverse. This is open for further investigation, as is the validation of the conclusion that Oregon's order estimates are higher.

5. CONCLUSIONS

Though the specific methodology of the studies summarized here differs in details from that of other regional studies (e.g. [6]), the general conclusion seems warranted that use of simple, or first order, Markov chain models of daily precipitation occurrence is sufficient for such objectives as estimations of the probabilities of spells and sequences of various kinds. There is evidence in our study, however, that in some areas—tentatively characterized as being climatologically diverse within a season—orders greater than 1 may be more appropriate. The reference of Feyerherm and Bark [7] to a tendency of this sort in April in the upper Midwest adds to the credibility of this conclusion.

A second conclusion is that threshold values separating wet days from dry days should be relatively small if a simple chain is used. If thresholds are above, say, 0.10 in. a day, the resulting analysis probably includes as "dry" too many days which are clearly "wet" in terms of their dynamic, or "air mass" potentialities for production of precipitation. The result is likely to be a process which tests as a random process: order 0. In this case the Chi-square methods proposed are unable to distinguish between randomness and the lack of suitably long records. This comment suggests that future research might profitably investigate models employing three, and where record length permits, four states.

REFERENCES

1. G. M. Adotevi-Akue, "Evaluation of Alternative Markovian Models for Precipitation Occurrence in Oregon," M.S. thesis, Oregon State University, Corvallis, June 1965, 29 pp.
2. L. Besson, "On the Probability of Rain," *Monthly Weather Review*, Vol. 52, No. 6, June 1924, p. 308.
3. P. Billingsley, "Statistical Methods in Markov Chains," *Annals of Mathematical Statistics*, Vol. 32, No. 1, Institute of Mathematical Statistics, Mar. 1961, pp. 12–40.
4. C. E. P. Brooks and N. Carruthers, *Handbook of Statistical Methods in Meteorology*, Her Majesty's Stationary Office, London, 1953, 413 pp.
5. J. E. Caskey, Jr., "A Markov Chain Model for the Probability of Precipitation Occurrence in Intervals of Various Length," *Monthly Weather Review*, Vol. 91, No. 6, June 1963, pp. 298–301.
6. A. M. Feyerherm and L. D. Bark, "Statistical Methods for Persistent Precipitation Patterns," *Journal of Applied Meteorology*, Vol. 4, No. 3, June 1965, pp. 320–328.
7. A. M. Feyerherm and L. D. Bark, "Goodness of Fit of a Markov Chain Model for Sequences of Wet and Dry Days," *Journal of Applied Meteorology*, Vol. 6, No. 5, Oct. 1967, pp. 770–773.
8. M. Finkner, "Progress Report No. 5, New Mexico Contributing Project to the Western Regional Research Project W-48," New Mexico Agricultural Experiment Station, Las Cruces, N. Mex., Sept. 1967 (unpublished manuscript).
9. K. R. Gabriel and J. Neumann, "A Markov Chain Model for Daily Rainfall Occurrence at Tel Aviv," *Quarterly Journal of the Royal Meteorological Society*, Vol. 88, No. 375, Jan. 1962, pp. 90–95.
10. J. R. Green, "A Model for Rainfall Occurrence," *Journal of the Royal Statistical Society*, Ser. B, Vol. 26, No. 2, London, 1964, pp. 345–353.
11. D. Heermann, "Evaluation of Markovian Models for Precipitation Occurrence in Colorado," Dept. of Agricultural Engineering, Colorado State University, Fort Collins, 1966, 13 pp. (unpublished manuscript).
12. P. G. Hoel, "A Test for Markoff Chains," *Biometrika*, Vol. 41, Parts 3 and 4, University Press, Cambridge, Dec. 1954, pp. 430–433.
13. J. W. Hopkins and P. Robillard, "Some Statistics of Daily Rainfall Occurrence From the Canadian Prairie Provinces," *Journal of Applied Meteorology*, Vol. 3, No. 5, Oct. 1964, pp. 600–602.
14. L. L. Weiss, "Sequences of Wet or Dry Days Described by a Markov Chain Probability Model," *Monthly Weather Review*, Vol. 92, No. 4, Apr. 1964, pp. 169–176.

[*Received February 9, 1968; revised April 17, 1968*]